Original Paper

Chronic Respiratory Disease

Chronic Respiratory Disease
Volume 16: 1–8
© The Author(s) 2018
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1479972318809452
journals.sagepub.com/home/crd
SAGE

# Cluster analysis identifying patients with COPD at high risk of 2-year all-cause mortality

**Antenor Rodrigues[1,2], Carlos Augusto Camillo[1], Karina Couto Furlanetto[1,3], Thais Paes[1], Andrea Akemi Morita[1], Thamyres Spositon[1], Leila Donaria[1,3], Marcos Ribeiro[1,4], Vanessa Suziane Probst[5], Nidia Aparecida Hernandes[1] and Fabio Pitta[1]**

## Abstract

The objective of the article is to identify clusters of patients with COPD according to factors known to be associated with mortality and to verify whether clusters' assignment is associated with 2-year mortality. Patients ($n = 141$) were evaluated by bioelectrical impedance, maximal inspiratory pressure (MIP), one-repetition maximum test of the quadriceps femoris (1RMQF) and BODE index (body mass index; airflow obstruction (spirometry); dyspnea (modified Medical Research Council scale); and exercise capacity (6-minute walk test (6MWT) distance). Vital status was retrospectively checked 2 years after the assessments, and time to death was quantified for those deceased in this period. *K*-means analysis identified two clusters. Patients in cluster one (CL I, $n = 69$) presented an impaired clinical status in comparison to cluster two (CL II, $n = 72$). Receiver operating characteristics curves identified the cutoffs discriminating patients composing CL I: forced expiratory volume in the first second <44%pred; 6MWT <479 m; 1RMQF <19 kg; and maximum inspiratory pressures <73 $cmH_2O$ (area under the curve range 0.750–0.857). During the follow-up, 19 (13%) patients deceased, 15 in CL I (22%) and 4 in CL II (0.06%) ($p = 0.005$). CL I was associated with a higher risk of 2-year mortality (hazard ratio (95% confidence interval): 4.3 (1.40–12.9), $p = 0.01$). A cluster of patients with COPD highly associated with 2-year mortality was statistically identified, and cutoffs to identify these subjects were provided.

## Keywords

Pulmonary disease, chronic obstructive, mortality

Date received: 6 November 2017; accepted: 18 September 2018

## Introduction

Chronic obstructive pulmonary disease (COPD)[1] was accountable for more than three million deaths in 2002, representing 5% of all deaths worldwide.[2] Worrisome, despite the amount of financial investments[3] and research and development, it is estimated to be the third leading cause of death by 2030.[2] Regardless of having its diagnosis based on lung function measurements, extrapulmonary manifestations are

[1] Laboratory of Research in Respiratory Physiotherapy (LFIP), Universidade Estadual de Londrina (UEL), Londrina, Brazil
[2] Departament of Physiotheraphy, Faculdade Pitágoras de Londrina, Londrina, Paraná, Brazil
[3] Center of Health and Biological Sciences, Universidade do Norte do Paraná, Londrina, Brazil
[4] Section of Pulmonology, Department of Medicine, Health Science Centre, Universidade Estadual de Londrina, Londrina, Brazil
[5] Center for Research and Post-Graduation in Health Sciences (CEPPOS), Health Sciences Center, Universidade Estadual de Londrina (UEL), Londrina, Brazil

**Corresponding author:**
Fabio Pitta, Laboratório de Pesquisa em Fisioterapia Pulmonar (LFIP), Departamento de Fisioterapia, Universidade Estadual de Londrina (UEL), Avenida Robert Koch, 60 – Vila Operária, 86038-350, Londrina – Paraná, Brazil.
Email: fabiopitta@uol.com.br

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (http://www.creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).

undoubtedly involved in disease progression and prognosis.[1,4] Within this context, a substantial scientific effort has been devoted to understand the factors involved in disease's association with the ultimate outcome in health care, that is, mortality (Online Supplemental Material, Table S1).[5–13]

Single variables were shown to be associated with mortality in COPD,[6–9,12] and multidimensional indices have also been developed intending to increase the capacity to identify patients at high risk of death.[5,10,11] Noteworthy, one of those indices which is based on the body mass index (BMI), airflow obstruction (forced expiratory volume in the first second ($FEV_1$)), dyspnea (modified Medical Research Council scale (mMRC)), and exercise capacity (6-minute walk test (6MWT) distance), the BODE index,[5] has been widely used and recognized. Despite the fair capacity of these multidimensional indices to predict survival, their composition include outcomes based on expert opinion,[9] fact that is known for not providing the best scientific evidence level,[14] since expert opinions are, at least in part, influenced by the clinical characteristics of the patients routinely seen in practice. Thus, between-location differences in patients' characteristics could be a determinant of the expert opinion and also of the accuracy of any index (or even a single variable) in predicting prognosis.

The variables composing the BODE index and each respective cutoff were identified in a cohort of 207 patients and further validated in a distinct cohort of 625 COPD patients from Spain, Venezuela, and the United States. Noteworthy, differences across geographical locations are noted for clinical characteristics in the cohort such as exercise capacity (i.e. 6MWT distance: $225 \pm 40$ in Venezuela, $311 \pm 121$ in the United States, and $446 \pm 99$ in Spain; $p < 0.001$). Although this variability guarantees that between-location patient characteristics are taken into account, proper validation is required for extrapolation of the results (i.e. predicting responses to values outside the range of the predictor variables used in the model) in samples with characteristics out of the patterns from the populations included in any cohort.

Notwithstanding, this is the case of Brazilian patients with COPD, who have already been shown to have higher 6MWT distance than patients from Europe or North America (i.e. a shift in the regression line curve to the left). Hence, it was hypothesized that by using a statistical method with the ability to identify clusters of patients according to well-known literature-based predictors of mortality previously identified in COPD, this would improve the capacity to identify the patients at high risk of mortality. This would allow identifying patients characterized by different combinations of risk factors, instead of identifying a patient with worse/better prognosis according to each single variable and summing their single weights.

Thus, to prove this hypothesis, this study aimed to identify clusters of patients with COPD according to factors previously known to be associated with mortality, verifying whether these clusters' assignment is associated with 2-year mortality; and secondarily, to compare the association of these clusters versus the association of the most recognized multidimensional index in COPD, i.e. the BODE index.[5]

## Materials and methods

### Study participants

For this retrospective cohort study with an intention-to-treat analysis (see statistical analysis for more information), patients with diagnosis of COPD according to the Global Initiative for Chronic Obstructive Lung Disease (GOLD),[1] who were referred to our institutional exercise training program and had data regarding mortality available at our database, were included in this study. Whether or not any of the patients was evaluated more than once, only the first assessment was used for further analysis. Inclusion criteria were those for screening to the exercise program, that is, clinical stability (i.e. absence of infections or exacerbations) in the previous 3 months, absence of severe and/or unstable cardiac diseases, and not having severe neurological or musculoskeletal conditions that might interfere in the proposed protocol. Patients unable to perform the proposed assessments were excluded. For database and statistical analyses handling, patients were unnamed and identified by a unique study code. Approval for study execution was obtained from the ethics committee of the State University of Londrina, Brazil, approval 996.413, and all patients signed a written informed consent.

### Vital status

The Center for Information on Mortality (CIM), the agency responsible for registering all death events at Londrina, Brazil, was used for official date of death confirmation for the deceased patients. Formal

authorization for using the CIM database was obtained from the Health Secretary of the city. Therefore, vital status was last checked, retrospectively, in January 2017, and quantified as time (i.e. months) from the last assessment to date of death for those who deceased. For those who were not at the CIM database, telephone contact was done to guarantee whether they were alive, and time to last contact was quantified. The date of the preceding contact was used for quantification of the follow-up for those patients who were not located, neither by telephone contact nor at the CIM database, at the last attempt.

### Measurements

A comprehensive initial questionnaire was used for characterization of the study sample, including information regarding anthropometrics, clinical history, and comorbidities. Lung function was assessed by spirometry (Spirobank spirometer, version 3.6 MIR, Rome, Italy),[15] inspiratory muscle strength by maximum inspiratory pressures (MIP; manovacuometry (Makil, Brazil)),[16] and functional exercise capacity by the 6MWT.[17] All the above cited tests followed specific international protocols and local reference values were used.[18–20] Body composition was assessed by bioelectrical impedance.[21] Peripheral muscle strength was measured by the one-repetition maximum test of the quadriceps femoris,[22] and dyspnea sensation in daily life by the mMRC.[23] Lastly, the BODE index (i.e. BMI, airflow obstruction ($FEV_1$), dyspnea (mMRC scale), and exercise capacity (6MWT distance)) was calculated.[5]

### Statistical analysis

The Statistical Package of Social Science (SPSS) 20 (SPSS Inc., Chicago, Illinois, USA) was used for statistical analysis. Normality in continuous data distribution was analyzed by the Shapiro–Wilk test. Continuous data following Gaussian distribution were expressed as means $\pm$ standard deviation. Otherwise, non-Gaussian distributed continuous data were expressed as median (interquartile range 25–75%]. For categorical data, the number of cases and its respective percentages were chosen for data expression. Furthermore, previously known predictors of mortality in patients with COPD were identified by a non-systematic literature review. Among the identified factors, those that were available at our database (i.e. $FEV_1$ (%predicted),[9] 6MWT (m),[7] fat-free mass index (kg/m$^2$),[6] quadriceps femoris strength (kg),[8] mMRC scale [0–4 points],[12] and maximum inspiratory pressure (cm$H_2O$)[13]; see Online Supplemental Information Table S1 for further information) were used to identify clusters of patients by *K*-means cluster analyses after data standardization. Additionally, the receiver operating characteristics (ROC) curve was performed to identify which of these variables are able to discriminate patients composing each cluster; an area under the curve (AUC) >0.500 was considered as the statistical cutoff to assume that a unique variable has discriminatory capacity to identifying patients composing each cluster. The Youden's index was used to identify the cutoffs associated with better specificity and sensitivity. As patients were evaluated at the intake of an exercise training program, an intention-to-treat analysis was carried out, by the last observation carried forward method, for those patients who did not finish the exercise training program; for those patients who finished the program, the post-exercise training measurements were considered. A posteriori, clusters' validation was done by verifying its association with 2-year mortality by Kaplan–Meier approach with log-rank test. Cox's proportional-hazard regression was used for adjust estimates of potential confounders (i.e. BODE index, change in 6MWT (m) after exercise training, gender, comorbidities, and pack years). Noteworthy, variables used to identify the clusters were not considered as confounders, as they are inherent to clusters composition. Non-paired *t*-test and $\chi^2$ test were also used when appropriate. A $p$ value lower than 0.05 was adopted.

## Results

Out of the 162 patients registered in our exercise training database, 153 had available data regarding mortality. Twelve of these 153 patients were not able to be clustered because of missing data of at least one of the factors included in the *K*-means model. Accordingly, a sample of 141 COPD patients was used in the final analysis, mostly classified as having moderate to very severe COPD. Entire characteristics' description of the sample is shown in Table 1. *K*-means cluster analysis was able to identify two clusters of patients, with 69 patients composing cluster one (CL I) and 72 cluster two (CL II). CL I patients had a more affected clinical status, having worse lung function, dyspnea sensation in daily life and exercise capacity, lower peripheral muscle strength, and fat-free mass index (Table 2). Further analysis revealed all variables used for cluster composition as having powerful

**Table 1.** Baseline characteristics of the study participants.[a]

| | $n = 141$ |
|---|---|
| Age (years) | 70 ± 8 |
| Gender, *n* (♀/♂) | 62/79 |
| BMI (kg/m$^2$) | 24.6 [21–29] |
| FVC (l) | 2.29 ± 0.78 |
| FVC (%pred) | 66 ± 17 |
| $FEV_1$ (l) | 1.04 [0.70–1.55] |
| $FEV_1$ (%pred) | 42 ± 17 |
| $FEV_1$/FVC | 49 [36–65] |
| GOLD, I/II/III/IV | 1/52/57/31 |
| FFMI (kg/m$^2$) | 16.2 ± 2.3 |
| 6MWT (m) | 477 [424–531] |
| 6MWT (%pred) | 88 [79–99] |
| 1RMQF (kg) | 18 ± 8 |
| mMRC (0–4 pts) | 3 [1–3] |
| MIP (cm$H_2O$) | 69 ± 23 |
| MIP (%pred) | 76 ± 25 |
| BODE index (I/II/III/IV) | 1 [1–1] |
| Comorbidities (*n*) | 2 [1–4] |
| Smoking history (pack-years) | 46 [18–68] |
| Follow-up (months) | 24 [24–24] |

CL I: cluster one; CL II: cluster two; ♀: female; ♂: male; BMI: body mass index; FVC: forced vital capacity; $FEV_1$: forced expiratory volume in the first second; GOLD: global initiative for chronic obstructive pulmonary disease assignment; FFMI: fat-free mass index; 6MWT: 6-minute walk test; 1RMQF: one-repetition maximum test of quadriceps femoris; mMRC: modified medical research council scale; MIP: maximum inspiratory pressure; BODE index: BMI, airflow obstruction ($FEV_1$); dyspnea (MRC), and exercise capacity (6MWT).

[a]Values are shown as mean ± standard deviation, median (interquartile 25–75%) or number of cases.

discriminatory capacity to identify patients with higher likelihood of being assigned to CL I (Table 3).

The follow-up period varied between 1 month and 24 months. Within this period, 19 (13%) patients deceased, 15 in CL I (22%) and 4 in CL II (0.06%) ($p = 0.005$). The Kaplan–Meier approach and univariate Cox's proportional-hazard regression analysis unveiled a statistically significant association between clusters' assignment and 2-year mortality, with patients assigned to CL I presenting more than four times increased chance to decease in 2 years (Table 4; Figure 1(a)). Noteworthy, in our sample, BODE index was not associated with survival (Table 4; Figure 1(b)). After correction for possible confounders in the multivariate model, clusters' assignment and gender were the only variables that remained in the model, with patients in CL I having more than 5.17 times increased chance to decease (Table 4).

## Discussion

Based on previously known predictors of mortality in COPD, two clusters were statistically identified, and further use of survival models unveil an outstanding association with 2-year mortality. Remarkably, the use of the BODE index as a comparator highlighted the promising potential of the present results for this population, since the BODE is the most scientifically used and recognized multidimensional prognostic index in COPD. However, it is anticipated that validation of these clusters in different samples involving worldwide population-based studies is necessary before its full comprehension and acceptance. Thresholds enabling identification of patients with higher likelihood of being assigned to CL I are also useful due to the fact that, for example, the BODE index was not able to identify those patients from our cohort who deceased since they were characterized by less impaired/better preserved exercise capacity, an important component of the BODE. Therefore, these thresholds might prove themselves as valuable alternatives to identify patients with a higher likelihood of composing CL I, and consequently increased chance to decease, in populations presenting similar characteristics. Altogether, these findings make these results a promising tool for the evaluation of "short-term" mortality risk in patients with COPD.

Multidimensional indices have already proved to be better than a sole variable in identifying patients at high risk of mortality, independently whether derived from measurements of lung function or from extrapulmonary manifestations of the disease.[5,10,11] In this study, the use of previously known predictors of mortality in COPD follows the same premise. The novelty, herein, is the use of an advanced statistical method to cluster patients. The identified clusters presented a multitude of differences, with patients in CL I considerably more affected than those from CL II and establishing a between-clusters known-groups validity (Table 2). This finding is in line with the original study proposing the BODE index, in which patients with worse scores were also more affected.[5] However, clusters' assignment in the present study was highly associated with mortality but not the BODE index (Table 4; Figure 1). In fact, a short follow-up was investigated in the present study, instead of the relatively longer follow-up investigated by Celli and colleagues[5] as the main outcome (24 vs. 52 months). This may, at least in part, help in the explanation of this finding. Furthermore, patients in

**Table 2.** Baseline characteristics and comparisons of the two identified clusters.[a]

| | CL I ($n = 69$) | CL II ($n = 72$) | $p$ |
|---|---|---|---|
| Age (years) | 68 ± 8 | 64 ± 8 | 0.02 |
| Gender, $n$ (♀/♂) | 36/33 | 26/46 | 0.06 |
| BMI (kg/m$^2$) | 23.1 [19.6–27.2] | 28 [23.8–31.6] | <0.001 |
| FVC (l) | 1.8 [1.4–2.2] | 2.5 [2–3] | <0.001 |
| FVC (%pred) | 58.6 [49–64.8] | 70 [62.5–84.8] | <0.001 |
| $FEV_1$ (l) | 0.78 [0.65–1] | 1.5 [1.1–1.8] | <0.001 |
| $FEV_1$ (%pred) | 33 ± 10 | 54 ± 15 | <0.001 |
| $FEV_1$/FVC | 43 [35–55] | 62 [51–67] | <0.001 |
| GOLD, I/II/III/IV | 0/8/36/25 | 1/447/21/6 | 0.000 |
| FFMI (kg/m$^2$) | 15 .4 ± 2.3 | 17.3 ± 1.8 | <0.001 |
| 6MWT (m) | 435 [368–490] | 524 [480–562] | <0.001 |
| 6MWT (%pred) | 80 [68–88] | 96 [87–103] | <0.001 |
| 1RMQF (kg) | 13.8 ± 6.1 | 21.2 ± 8.1 | <0.001 |
| mMRC (0–4 pts) | 3 [3–3] | 2 [1–3] | <0.001 |
| MIP (cm$H_2O$) | 57 ± 18 | 79 ± 23 | <0.001 |
| MIP (%pred) | 66 ± 20 | 86 ± 26 | <0.001 |
| BODE index, I/II/III/IV | 1 [1–5] | 1 [1–1] | <0.001 |
| Comorbidities ($n$) | 3 [2–4] | 2 [1–3] | 0.08 |
| Smoking history (pack-years) | 42 [16–68] | 42 [21–61] | 0.80 |
| Follow-up (months) | 24 [20–24] | 24 [24–24] | 0.09 |

CL I: cluster one; CL II: cluster two; ♀: female; ♂: male; BMI: body mass index; FVC: forced vital capacity; $FEV_1$: forced expiratory volume in the first second; GOLD: global initiative for chronic obstructive pulmonary disease assignment; FFMI: fat-free mass index; 6MWT: six-minute walk test; 1RMQF: one-repetition maximum test of quadriceps femoris; mMRC: modified medical research council scale; MIP: maximum inspiratory pressure; BODE index: BMI, airflow obstruction ($FEV_1$); dyspnea (MRC), and exercise capacity (6MWT).
[a]Values are shown as mean ± standard deviation, median (interquartile 25–75%) or number of cases.

**Table 3.** ROC curve analyses identifying patients composing CL II.

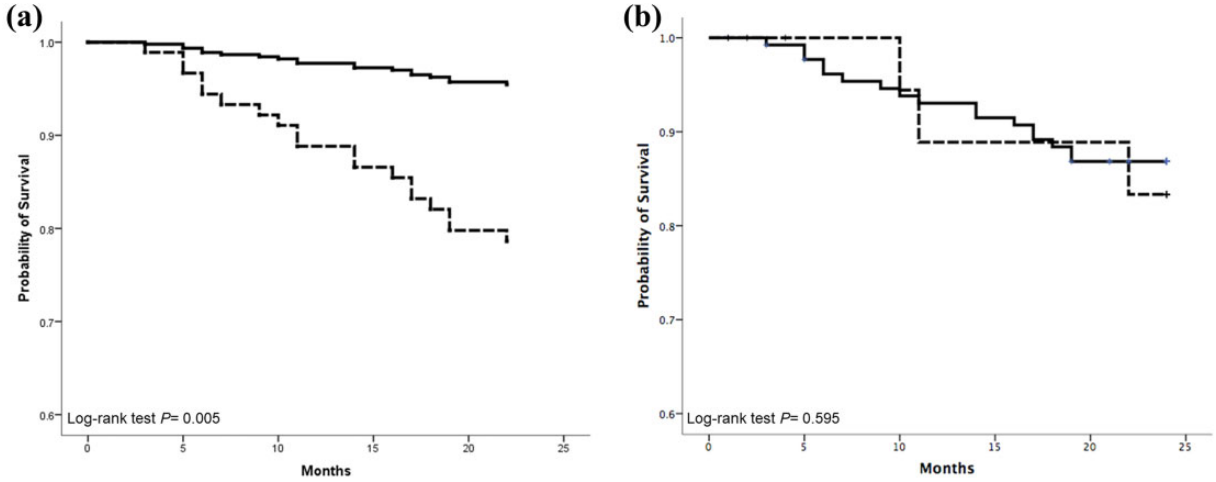
| | $FEV_1$ (%pred) | 6MWT (m) | FFMI | QF (kg) | MIP (cm$H_2O$) | mMRC |
|---|---|---|---|---|---|---|
| Cutoff | 44 | 479 | 16.7 | 18.5 | 73 | 3 |
| Sensitivity | 0.841 | 0.739 | 0.812 | 0.797 | 0.754 | 0.783 |
| Specificity | 0.764 | 0.764 | 0.653 | 0.667 | 0.681 | 0.653 |
| AUC | 0.857 | 0.796 | 0.750 | 0.768 | 0.784 | 0.782 |

CL II; cluster two; 6MWT: 6-minute walk test; $FEV_1$: forced expiratory volume in first second; QF: quadriceps femoris strength; MIP: maximum inspiratory pressure; mMRC: modified medical research; FFMI: fat-free mass index; AUC: area under the curve.

**Table 4.** Risk of 2-year mortality due to any cause.

| | Univariate | | Multivariate[a] | |
|---|---|---|---|---|
| | Hazard ratio (95% CI) | $p$ Value | Hazard ratio (95% CI) | $p$ Value |
| Cluster I versus II | 4.3 (1.4–12.9) | 0.01 | 5.17 (1.7–15.7) | 0.004 |
| BODE index (points) | 1.13 (0.92–1.39) | 0.24 | — | NS |
| Δ6MWT (m) | 0.1 (0.98–1) | 0.07 | — | NS |
| Gender | 2.95 (0.98–8.88) | 0.06 | 3.77 (1.24 – 11.4) | 0.019 |
| Comorbidities | 1.2 (0.79–5.07) | 0.15 | — | NS |
| Smoking history (pack years) | 0.1 (0.98–1.02) | 0.95 | — | NS |

95% CI: 95% confidence interval; BODE index: BMI, airflow obstruction ($FEV_1$); dyspnea (MRC), and exercise capacity (6MWT); Δ6MWT: change in the 6MWT distance after exercise training; Comorbidities: patients with more than two comorbidities versus those with less than two; NS: non-statistically significant; mMRC: modified medical research.
[a]Variables with a $p$ value lower than 0.20 were considered as significant for adjustment at the multivariate level.




| | 0-6 months | 6-12 months | 12-18 months | 18-24 months | 24 months |
|---|---|---|---|---|---|
| CL 1, n/death events/censured | 69/3/3 | 63/5/0 | 58/4/0 | 54/3/0 | 51/0/- |
| CL 2, n/death events/censured | 72/0/3 | 69/2/0 | 67/1/0 | 66/1/4 | 61/0/- |
| BODE index I/II, n/death events/censured | 122/3/3 | 116/5/0 | 111/5/0 | 106/3/4 | 99/0/- |
| BODE index III/IV, n/death events/censured | 19/0/3 | 16/2/0 | 14/0/0 | 14/1/0 | 13/0/- |

**Figure 1.** Life table and Kaplan–Meier survival curves with logrank test. (a) For the identified clusters (shorter dashed plot indicate CL I patients; longer dashed plot indicate CL II patients); (b) For the BODE index (longer dashed plot indicate—BODE I/II patients; shorter dashed plot indicate BODE III/IV patients) during the 24-month follow-up period.

the present cohort had a better preserved exercise capacity (Table 1), and even patients who deceased in the follow-up period walked longer distances in the 6MWT than the patients who did not decease in the BODE cohort[5] (mean 6MWT distance (m) of 384 $\pm$ 101 vs. 264 $\pm$ 113, respectively). Corroborating the role of between-location differences in patients' characteristics, only 17 (12%) patients in our sample walked less than 350, the highest cutoff for the 6MWT in the BODE index. This is not surprising, since it is already known that socioeducational–cultural differences (and consequently geographical location) may influence patients' level of exercise capacity.[24] Therefore, these differences should be taken into consideration when interpreting the present results.

Despite the abovementioned differences in study design and sample, the survival curve of the BODE index noticeably highlights patients in the "better" (or "milder") assignments (i.e. BODE I/II) having a steep death rate until approximately 10 months of follow-up (i.e. the survival curves cross approximately at this time point; Figure 1(b)). This indicated patients with a better clinical status according to the BODE index deceasing at shorter time and also justified the higher 6MWT distance cutoff identifying patients composing CL I (i.e. 479 m; Table 3) in comparison to the previously published value of 350 m (Online Supplemental Material, Table S1).[7] Additionally, two other reasons may help explaining the fact that the present clusters' assignment achieved better short-term mortality associations than the BODE index: the substitution of variables/cutoffs proposed by expert opinions by using a statistical analysis method to identity clusters of patients from variables previously known as predictors of mortality; and the variables with the ability to identify the clusters (Table 3) having a bigger impact in short-term mortality than the variables used to classify patients according to the BODE index[5] (i.e. BMI, $FEV_1$, mMRC, and 6MWT).

Despite the fact that an increased chance to decease in patients presenting worse clinical status is not surprising, finding ways to identify these patients, however, is an important clinical information. Thus, the use of these simple and easily applicable cutoffs presented in Table 3 would enable the identification of patients with higher likelihood of short-term mortality. The level of obstruction, measured by the $FEV_1$, was the cutoff with higher discriminatory capacity (AUC 0.857); 6MWT, MIP, dyspnea in daily life, peripheral muscle strength, and body composition also presented powerful discriminatory capacity to

identifying patients composing CL I (AUC 0.750–0.796). Therefore, simple measurements, routinely performed in clinical practice,[4] can be used with this purpose. Indeed, clusters based on these aforementioned variables, all of which previously known to be associated with mortality,[6–9,12,13] help in explaining their higher discriminatory capacity. Moreover, as these are measurements routinely found in clinical practice, this easily allows to use any of these cutoffs, either independently or in combination. For those clinical situations in which more than one, or all of these variables are available, patients could be classified as having smaller or higher likelihood of short-term mortality according to the number of criteria met. As an example, patients presenting values below the cutoff of six variables may have a higher likelihood to decease than patients presenting values below the cutoff of only one variable, though external validation is needed to support or refute this hypothesis.

Despite all the efforts applied in the development of the present study, some limitations should be pointed out. The retrospective design is known for not being the preferred design for cohort studies. To circumvent this limitation, corrections for comorbidities, improvement in 6MWT after the exercise training, and other possible confounders were done (Table 4). However, other factors, such as the occurrence of acute exacerbations during the follow up period, were not possible to be adjusted for, as this information was not available in our database. Also, selection bias cannot be ruled out as a limitation due to the fact that the screening for inclusion was done at the initial assessment of a rehabilitation program. On the other hand, nowadays, not offering exercise training for patients with COPD could even be considered unethical. Noteworthy, disease severity did not preclude the offer to take part in the exercise training program and, consequently, participation in the present study. Moreover, to circumvent the influence of exercise training in patients' survival, improvement in 6MWT after exercise training was used as a confounder in the analysis (Table 4).

In summary, two clusters of patients were identified using variables previously known as predictors of mortality in COPD, and clusters' assignment was proved highly associated with 2-year mortality. The cutoffs allowing the identification of these clusters were provided and are easily applicable in clinical practice. However, further validation of these clusters and cutoffs in different samples worldwide is welcomed to endorse or refute their applicability.

## Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The study was partially funded by a personal grant from the Coordination for the Improvement of Higher Education Personnel (CAPES), Brazil.

## ORCID iD

Fabio Pitta https://orcid.org/0000-0002-3369-6660

## Supplemental material

Supplemental material for this article is available online.

## References

1. Vogelmeier CF, Criner GJ, Martinez FJ, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive lung disease 2017 report. GOLD executive summary. *Am J Respir Crit Care Med* 2017; 195(5): 557–582.
2. World Health Organization. Chronic respiratory disease: Burden of COPD,http://www.who.int/respiratory/copd/burden/en/ (2017, accessed 25 April 2017).
3. Ford ES, Murphy LB, Khavjou O, et al. Total and state-specific medical and absenteeism costs of COPD among adults aged >/= 18 years in the United States for 2010 and projections through 2020. *Chest* 2015; 147(1): 31–45.
4. Spruit MA, Singh SJ, Garvey C, et al. An official American thoracic society/European respiratory society statement: key concepts and advances in pulmonary rehabilitation. *Am J Respir Crit Care Med* 2013; 188(8): e13–e64.
5. Celli BR, Cote CG, Marin JM, et al. The body-mass index, airflow obstruction, dyspnea, and exercise capacity index in chronic obstructive pulmonary disease. *N Engl J Med* 2004; 350(10): 1005–1012.
6. Vestbo J, Prescott E, Almdal T, et al. Body mass, fat-free body mass, and prognosis in patients with chronic obstructive pulmonary disease from a random population sample: findings from the copenhagen city heart study. *Am J Respir Crit Care Med* 2006; 173(1): 79–83.
7. Cote CG, Pinto Plata V, Kasprzyk K, et al. The 6-min walk distance, peak oxygen uptake, and mortality in COPD. *Chest* 2007; 132(6): 1778–1785.
8. Swallow EB, Reyes D, Hopkinson NS, et al. Quadriceps strength predicts mortality in patients with

moderate to severe chronic obstructive pulmonary disease. *Thorax* 2007; 62(2): 115–120.
9. Almagro P, Martinez Camblor P, Soriano JB, et al. Finding the best thresholds of FEV1 and dyspnea to predict 5-year survival in COPD patients: the COCOMICS study. *PLoS One* 2014; 9(2): e89866.
10. Puhan MA, Garcia-Aymerich J, Frey M, et al. Expansion of the prognostic assessment of patients with chronic obstructive pulmonary disease: the updated BODE index and the ADO index. *Lancet* 2009; 374(9691): 704–711.
11. Puhan MA, Hansel NN, Sobradillo P, et al. Large-scale international validation of the ADO index in subjects with COPD: an individual subject data analysis of 10 cohorts. *BMJ Open* 2012; 2(6): e002152.
12. Casanova C, Marin JM, Martinez Gonzalez C, et al. Differential effect of modified medical research council dyspnea, COPD assessment test, and clinical COPD questionnaire for symptoms evaluation within the new GOLD staging and mortality in COPD. *Chest* 2015; 148(1): 159–168.
13. Gray Donald K, Gibbons L, Shapiro SH, et al. Nutritional status and mortality in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 1996; 153(3): 961–966.
14. Centre For Evidence-based Medicine. Oxford centre for evidence-based medicine-levels of evidence, http://www.cebm.net/oxford-centre-evidence-based-medicine-levels-evidence-march-2009/ (2009, accessed 25 April 2017).
15. Miller MR, Hankinson J, Brusasco V, et al. Standardisation of spirometry. *Eur Respir J* 2005; 26(2): 319–338.
16. Black LF and Hyatt RE. Maximal respiratory pressures: normal values and relationship to age and sex. *Am Rev Respir Dis* 1969; 99(5): 696–702.
17. ATS Committee on Proficiency Standards for Clinical Pulmonary Function Laboratories. ATS statement: guidelines for the six-minute walk test. *Am J Respir Crit Care Med* 2002; 166(1): 111–117.
18. Pereira CA, Sato T and Rodrigues SC. New reference values for forced spirometry in white adults in Brazil. *J Bras Pneumol* 2007; 33(4): 397–406.
19. Neder JA, Andreoni S, Lerario MC, et al. Reference values for lung function tests. II. Maximal respiratory pressures and voluntary ventilation. *Braz J Med Biol Res* 1999; 32(6): 719–727.
20. Britto RR, Probst VS, de Andrade AF, et al. Reference equations for the six-minute walk distance based on a Brazilian multicenter study. *Braz J Phys Ther* 2013; 17(6): 556–563.
21. Kyle UG, Pichard C, Rochat T, et al. New bioelectrical impedance formula for patients with respiratory insufficiency: comparison to dual-energy X-ray absorptiometry. *Eur Respir J* 1998; 12(4): 960–966.
22. Robles PG, Mathur S, Janaudis-Fereira T, et al. Measurement of peripheral muscle strength in individuals with chronic obstructive pulmonary disease: a systematic review. *J Cardiopulm Rehabil Prev* 2011; 31(1): 11–24.
23. Kovelis D, Segretti NO, Probst VS, et al. Validation of the modified pulmonary functional status and dyspnea questionnaire and the medical research council scale for use in Brazilian patients with chronic obstructive pulmonary disease. *J Bras Pneumol* 2008; 34(12): 1008–1018.
24. Pitta F, Breyer MK, Hernandes NA, et al. Comparison of daily physical activity between COPD patients from central Europe and South America. *Respir Med* 2009; 103(3): 421–426.